FHWA Workshop over the Web for Travel Model Development Session 2 Homework

Working with Survey Data

Total Time Required: 45-60 Minutes

Introduction

In this exercise, we will work with a real survey data set to accomplish three objectives:

1. Understand what is inside a household survey trip file;
2. Conduct some sample analysis on the survey data; and
3. Do some simple model estimation.

Part 1. Understanding what is inside a household survey trip file

Please download the MS Excel file "***Homework 2.xls***". You should see the trip data when you open this file. This is generally how most household survey trip files look. Please note that while we have provided data that have already been cleaned, most survey data are seldom as clean, and contain multiple missing values. Significant upfront effort is usually necessary to clean and assure the quality of the data.

The focus of this homework is not to learn how to clean the data, but rather to understand what types of analyses can be done with good data. In preparation for Parts 2 and 3, we will need to familiarize ourselves with the data in this worksheet.

- You will see 25 columns of data in the ***"Trips"*** worksheet. The descriptions of each of these fields is given in the worksheet named ***"Trip Data Dictionary".***
- Also included in the "***Trip Data Dictionary"*** worksheet are codes for the values of some fields. For example, when you see a value of 1 in the "Purpose" column of the "***Trips***" worksheet, you know that it is a "Journey-to-Work" trip, as per the data dictionary.
- Go through the various fields and their descriptions. Do you think there are any other columns that should be in the trip file? What other attributes of a person's trip record will be useful to you as a planner, modeler or consultant?

Part 2. Conduct some sample analysis on the survey data

Now that we are familiar with the data, it's time to do some fun analysis! Aside from being fun, data analysis is a critical step before any model is estimated. It gives us a good sense of the travel behavior patterns in the study region. ***How many work trips are there? How many public transit trips are there? Which sub-regions or districts produce the most number of trips and which ones attract the most?***

These are the kinds of questions we will be able to answer with data analysis. Needless to say, we could do lots and lots of analysis. Also, analyzing large data sets is often not very trivial and is usually done in statistical packages such as SAS or SPSS. But to keep things simple, we will stick with MS Excel and will analyze the distribution of trips by purpose and mode. Here are the steps you need to follow:

- Once again, we go back to our good friend "***Homework 2.xls***".
- This time click on the worksheet "***Part 2 Exercise".***
- You should see an empty grey table named ***Cross-Tabulation of Trips by Mode and Purpose.***
- Go ahead and click anywhere on this table. You will see a small menu called "PivotTable Field List" pop up to the right of the table. We will use this menu to populate our empty table.
- Using your mouse, scroll down the menu and select the field ***Mode.***
- Click on this item and drag it into the empty grey table to the location where you see the following message: ***"Drop Row Fields Here".***
- Now go back to the menu and scroll once again until you find the field ***"Purpose"***. Click, drag, and drop the "Purpose" field into the table to the location where you see the following message ***"Drop Column Fields Here"***.
- Now for the final part. Go back to the menu and scroll to the field named ***"WEIGHT"***. Drag and drop this item into the table where you see the following message ***"Drop Data Items Here".***
- Once you complete these steps, you will see that the tables named ***"Table 2. Percentage of Trips by Mode and Purpose"*** and "***Table 3. Percentage of Trips by Purpose"*** will be populated automatically.

<u>Questions for Homework</u>

- Use the results from Tables 2 and 3 to answer the following questions:
    - Which trip purpose contributes the highest number of trips to the region?
    - Which trip purpose shows the highest share of "Driver" mode trips and which purpose has the highest share of "Passenger" trips?

- Which are the top two trip purposes for public bus?

- Which purpose shows the highest percentage of non-motorized (walk and bike) trips.

Part 3. Model Estimation

For the last part of our homework, we will estimate a ***trip attraction*** model for Journey-to-Work trips. You will recall from the first webinar that the purpose of a model is to relate an ***dependent*** variable to an ***independent*** variable using a mathematical formula. Why do we need models? Well, once we know the relationship between a dependent variable and an independent variable, we will be able to forecast the dependent variable if we know the forecasts for the independent variable!

For Part 3, you will need to follow the steps below:

- Once again, go back to "***Homework 2.xls***".

- Click on the worksheet named ***"Part 3 Exercise".***

- You will see a simple table with three columns: attraction district name, number of work trip attractions and total employment in each district. This table has been created using the same pivot table techniques that we used in Part 2. But this time, for simplicity, we have directly provided you with the data themselves.

- Our goal will be to relate the Journey-to-Work attractions with the underlying total employment. This is one of the most common regressions in all transportation modeling. We want to find out ***how many trips will be attracted by each job in the region***. This is why we will estimate a ***regression model***.

- Before we proceed, you need to check if your Excel version allows Regression. In your Excel window, go to Tools -> Add-Ins. Make sure ***Analysis ToolPak*** is checked. Then click **OK.**

- Now go back to the worksheet ***Part 3 Exercis***e. Go to ***Tools->Data Analysis.*** Scroll down in the pop up window and select ***Regression.*** Then Click ***OK***

- You will now see a ***Regression*** input box. You should see that most of the entries are already filled in. If not, please do the following:

    - In the entry ***Input Y Range,*** type in $C$3:$C$27.

    - In the entry ***Input X range***, type in $D$3:$D$27.

    - In the entry ***Output Rang,*** type in $B$31

    - Check the entry ***Line Fit Plots***

    - Then click ***OK***

- You will notice that the model result shows up in cells ***B29 and C29***
- Phew! we are almost done. Now let us focus on interpreting the model.

Questions for Homework

- How do you interpret the trip attraction model?
- How many trips does each additional job generate?
- Does this number make sense?